Proceedings of the Eighteenth International Conference on
Civil, Structural and Environmental Engineering Computing
Edited by: P. Iványi, J. Kruis and B.H.V. Topping
Civil-Comp Conferences, Volume 10, Paper 16.2
Civil-Comp Press, Edinburgh, United Kingdom, 2025
ISSN: 2753-3239, doi: 10.4203/ccc.10.16.2

Integrating RAG with Visual Hazard Recognition for Automated Generation of Prevention Measures: A Preliminary Study

W.-D. Yu, W.-T. Hsiao and H.-H. Li

Department of Civil and Construction Engineering, Chaoyang University of Technology,
Taichung, Taiwan, R.O.C.

Abstract

Previous studies on construction site hazard identification have primarily employed machine learning techniques to predict potential accident scenarios. However, two major limitations persist: (1) these models often rely on manual input of attribute parameters, reducing their applicability in real-time and automated settings; and (2) most approaches focus solely on hazard prediction without providing specific, regulation-compliant preventive strategies. To address these gaps, this study proposes an automated framework that integrates computer vision with Retrieval-Augmented Generation (RAG) for hazard identification and response plan generation. Specifically, hazard types are detected from construction site CCTV footage using computer vision, while large language models (LLMs) are employed to retrieve relevant construction safety regulations and generate corresponding mitigation measures. Empirical validation was conducted using a dataset of 2,490 hazard images to test the proposed model. Results demonstrate that the LLM-RAG framework can generate feasible, regulation-aligned preventive recommendations. The model significantly enhances the automation and intelligence of hazard recognition and mitigation planning, offering a novel approach to advancing smart construction safety management.

Keywords: construction safety, hazard prevention, pre-trained models, image recognition, RAG, LLM.

1 Introduction

The construction industry, characterized by its open and dynamic work environments, has long been recognized as one of the most hazardous sectors globally [1]. According to the International Labour Organization (ILO), although construction employs only 7% of the global workforce, it accounts for 30–40% of all occupational fatalities [2]. Even in highly developed regions such as the European Union, the construction sector is responsible for approximately 20% of annual work-related deaths across all industries[3]. These statistics underscore the critical importance of enhancing occupational safety in construction.

A fundamental step in reducing occupational accidents is the prevention of workers' exposure to hazardous conditions [4]. Effective hazard prevention depends on the timely and accurate identification of potential hazards [5]. When hazards are not promptly recognized, they are unlikely to be addressed adequately, thereby increasing the risk of on-site accidents and injuries [6]. Consequently, hazard identification is widely regarded as a cornerstone of accident prevention strategies and an essential prerequisite for ensuring construction site safety.

However, previous studies have demonstrated the limitations of relying solely on workers' personal judgment to identify hazards. In the United States, it was found that over 40% of hazards were not recognized by workers [6]. Similar findings have been reported in Australia (57% unrecognized) [7]and the United Kingdom (33%) [8], with comparable trends observed across Asia, the Middle East, and other regions [9]. These findings suggest that the inability of workers to accurately identify latent hazards and take pre-emptive action is a major contributing factor to the persistently high accident and fatality rates in construction.

Several challenges have been identified in the literature as key obstacles to effective hazard recognition by construction workers: (1) Complex and dynamic site environments – construction sites are highly information-dense settings with changing layouts, multiple work phases, varied materials and equipment, and fluctuating environmental conditions, all of which hinder workers' ability to focus and identify specific hazards[10]; (2) Visual interference in imagery – site imagery often contains cluttered backgrounds, inconsistent lighting, and overlapping human activity, reducing observers' attention and accuracy in hazard recognition [11] (3) Cognitive load and individual differences – variations in workers' hazard perception abilities may limit the effectiveness of visual recognition tasks, particularly under high workload conditions [11]; (4) Subtlety and concealment of hazards – certain hazards (e.g., electrical leakage, structural defects) are not readily visible and require specialized knowledge or experience to detect [12][13].

In summary, the high complexity of construction environments and the cognitive limitations of human observers result in a significant proportion (33%–57%) of hazards being overlooked, which constitutes a key root cause of accidents in the construction sector [6][7][8][9].

Previous research has employed various machine learning algorithms—including Support Vector Machines (SVM) [14], Linear Regression (LR) [15], Random Forest (RF) [16], and Naive Bayes (NB) [17]—to predict construction-related accidents.

However, these models typically require manual input of hazard attribute data and are not capable of automatic hazard identification. More recently, some studies have leveraged computer vision and construction site CCTV footage to collect hazard data [18], but few have explored the integration of computer vision for automated hazard detection. Furthermore, while prior works have made progress in predicting potential accident types, they often fall short of proposing concrete and regulation-compliant preventive measures.

To address these limitations, this study proposes an integrated framework that combines computer vision-based hazard detection with Retrieval-Augmented Generation (RAG)[19] powered by large language models (LLMs)[20]. This framework enables automatic hazard recognition and generation of compliant preventive measures using CCTV surveillance footage from construction sites. By leveraging on-site cameras and local processing units, the proposed system facilitates 24/7 real-time monitoring, hazard detection, and safety response generation—offering a novel and effective approach to advancing construction site safety management.

2 Methods

The proposed model for identification and prevention of construction site hazards is depicted in Figure 1. There are two modules in the proposed model: (1) Deep Learning-based Hazard Prediction Module (DL-HPM) for identifying the construction hazard type; (2) Large Language Model-based Retrieval Augmented Generation Module (LLM-RAG) for generating the hazard prevention measures.

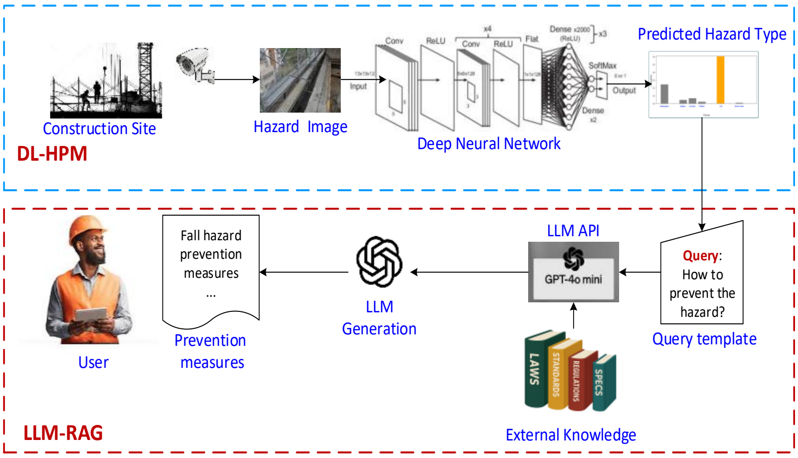

Figure 1: Proposed model for construction hazard identification and prevention.

2.1 Deep Learning-based Hazard Prediction Module (DL-HPM)

DL-HPM comprises of the following components: (1) The CCTV site safety monitoring system—providing 24/7 real-time safety monitoring footages of construction site; (2) LLM API—the Application Programming Interface (API) of large language model (LLM), ChatGPT o4 mini was adopted for this research; (3) Deep Neural Networks—a Convolutional Neural Newtok (CNN)-based deep network is adopted for pattern recognition of construction hazard on-site; (4) the most likely hazard type is predicted by the Deep Neural Networks as the outcome of DL-HPM.

2.2 Large Language Model-based Retrieval Augmented Generation Module (LLM-RAG)

LLM-RAG comprises of the following components: (1) Query template—providing a template of query to generate a description of request to query the hazard prevention measures for the hazard type predicted by DL-HPM; (2) External Knowledge—the domain-specific construction safety laws, standards, SPECs, and regulations published by the Occupational Safety and Health Administration, Taiwan (TOSHA); (3) LLM-API—the ChatGPT 4o mini is adopted to convert both external knowledge base and the hazard prevention query into vector embeddings; (4) LLM generation—ChatGPT 4o mini is adopted again to generate the solution for the query; (5) Hazard prevention report—a report of hazard prevention measures based on the construction safety laws, standards, SPECs, and regulations of TOSHA is generated for the user.

2.3 Performance Metrics

In order to evaluate the performance of the proposed methods, two sets of metrics were adopted in this research to evaluate the outcomes of the proposed model.

2.3.1 Performance Evaluation Metrics of Visual Hazard Recognition

The first set of metrics are frequently used for pattern recognitions, comprising Macro-Precision, Macro-Reall, Macro-F1 score, and Accuracy [23]:

$$Macro - Precision = \frac{1}{\mathrm{C}}\sum_{i=0}^{C}\frac{TP_i}{TP_{i+}FP_i} \tag{1}$$

where TP_i and FP_i denote the true positives and false positives for class i, and C is the number of classes.

$$Macro - Recall = \frac{1}{\mathrm{C}}\sum_{i=0}^{C}\frac{TP_i}{TP_{i+}FN_i} \tag{2}$$

where TP_i denotes the true positives, and FN_i denotes false negatives for class i, and C is the number of classes.

$$Macro - F1 = \frac{Macro-Precision \times Macro-Recall}{Macro-Precision + Macro-Recall} \tag{3}$$

where *Macro-Precision* and *Macro-Recall* are defined in Eq. (1) and Eq. (2), respectively..

$$Accuracy = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FP_i + FN + TN_i)} \tag{4}$$

where TP_i, FP_i,denote the true positives and false positives for class i; FN_i, TN_i,denote the false negatives and true negatives; and C is the number of classes.

2.3.2 Performance Evaluation Metrics of Hazard Prevention Measures

The second set of performance metrics evaluates the correctness of the prevention measures suggested by the proposed model. Established Natural Language Processing (NLP) evaluation metrics [24] include BLEU (n-gram precision against references), ROUGE (n-gram overlap/recall against references), CIDEr (similarity based on multi-reference consensus), and SPICE (semantic proposition comparison against references). However, these standard NLP metrics may not be suitable for evaluating the generated hazard prevention measures.

Since the proposed measures are derived and often paraphrased from TOSHA's construction safety laws, regulations, standards, and specifications, there are no absolute or definitive "correct" references available for direct comparison. Consequently, applying conventional evaluation metrics may lead to misleading conclusions. A more suitable approach is to assess the proposed prevention measures through expert evaluation by construction safety professionals who are certified by TOSHA and possess practical knowledge of the relevant regulatory documents. In this context, a Likert 5-point scale is adopted, and evaluations are conducted by an experienced domain expert certified in occupational safety and health (OSH) by TOSHA. Three key performance indicators are assessed: (1) Completeness of Hazard Description, (2) Appropriateness of the Recommended Prevention Measures, and (3) Correctness of the Source Documents. Each indicator is calculated using the following equation :

$$Mean\ Score = \frac{\sum_{i=1}^{N} Sore_i}{N} \tag{5}$$

where *Mean Score* can be 'Completeness of Hazard Description', 'Appropriateness of Proposed Prevention Measures', or 'Correctness of Referred Regulations' for the i^{th} case; and N is the number of cases.

3 Results

The proposed model has been tested with the CCTV footages collected from a real-world construction project [26] to recognize hazard types and to generated feasible prevention measures.

3.1 Performance Evaluation of Visual Hazard Recognition

For visual hazard recognition, we adopted a transfer learning approach utilizing pre-trained deep learning image recognition models: GoogleNet, Inception-v3, ResNet-101, and DenseNet-201. The model testing platform consisted of Matlab 2024b with the Computer Vision Toolbox, running on the following hardware: (1) CPU: Intel Xeon E5-2620v4 @ 2.10GHz; (2) RAM: 40GB at 2400MHz; (3) OS: Microsoft Windows 10; and (4) GPU: NVIDIA Quadro P2000 (5GB). We employed the Matlab® Deep Network Designer App to build the pre-trained models.

A set of 2,490 pre-labelled hazard images captured from the CCTV footages collected from the case project were used for model testing, where 80% (1,992 images) of the datasets were used for training and the rest 20% (498 images) for testing. Then, the four performance metrics described in Subsection 2.1 are used to evaluate the performance of the five pretrained models. The results are reported in Table 1.

Table 1 indicates that the visual hazard recognition performance for construction hazards is generally lower than that reported for other visual recognition tasks in the literature. This underperformance may stem from three primary factors: (1) the inherent complexity of construction scenes, which poses a challenge for common pre-trained models; (2) the small size (224×224×3) and low resolution of images captured from CCTV footage, potentially insufficient for large pre-trained models; and (3) the presence of compound hazards within the same scene, meaning a single image can depict multiple hazard types (as shown in the examples of Table 2).

Pre-train Model	Macro-Precision	Macro-Recall	Macro-F1	Accuracy
GoogleNet	0.6256	0.5753	0.5815	0.7664
Inception-v3	0.7010	0.5362	0.5622	0.7848
ResNet-101	0.7217	0.6446	0.6657	0.8189
Dese Net-201	0.7150	0.6254	0.6502	0.8084

Table 1: Visual recognition performance of pre-trained models.

No.	Image	Hazard description	Hazard type
1		Incorrectly used safety harnesses on scaffolding expose top-floor formwork workers to fall hazards (H01).	Fall & rolling (H01).
2		Materials piled on scaffolding walkways during top-floor formwork assembly create a tripping risk (H02).	Tripping (H02).

Table 2: Examples of compound hazards with similar scenes.

3.2 Performance Evaluation Metrics of Hazard Prevention Measures

3.2.1 External Knowledge Base for Construction Hazard Prevention

The construction safety regulations adopted for LLM-RAG modules including the 10 documents published by TOSHA of Taiwan, list as follows (The full articles and contents of these documents are available on the TOSHA website: https://www.osha.gov.tw/48110/48713/normalnodelist.):

1. Occupational Safety and Health Act (OSHA,職業安全衛生法, amended on 2019.05.05)
2. Enforcement Rules of the Occupational Safety and Health Act (EROSHA,職業安全衛生法施行細則, amended on 2020.02.27)
3. Occupational Safety and Health Management Regulations (OSHMR,職業安全衛生管理辦法, amended on 2016.02.19)
4. Occupational Safety and Health Facility Regulations (OSHFR,職業安全衛生設施規則, amended on 2020.03.02)
5. Regulations for Occupational Safety and Health Labeling (ROSHL,職業安全衛生標示設置準則, amended on 2014.07.02)

6. Notice for Labor Inspection Agencies Handling Category D Hazardous Work Environment Reviews(NLAHCD,勞動檢查機構辦理丁類危險性工作場所審查注意事項, revised Sept. 21, 2025)
7. Construction Safety and Health Facilities Standards (CSHFS,營造安全衛生設施標準, revised on June 26, 2014)
8. Labor Inspection Act (LIA,勞動檢查法, amended on 2015.02.04)
9. Standards for Determining Whether Workers are in Immediate Danger as Prescribed in Article 28 of the Labor Inspection Act (SDWID,勞動檢查法第二十八條所定勞工有立即發生危險之虞認定標準, revised on June 10, 2005)
10. Key Points for Strengthening Labor Safety and Health Management in Public Works (KPSLSH, 加強公共工程勞工安全衛生管理作業要, revised on December 2, 2009)

3.2.2 Calling ChatGPT 4o mini Model via API for Safety Hazard Prevention

In this research, we implemented the LLM-RAG model using ChatGPT-4o mini. The external knowledge base comprised the 10 TOSHA documents detailed in the preceding subsection. We developed an API program to interface with ChatGPT-4o mini. There two reason for choosing ChatGPT-4o mini for implementation: (1) reasoning quality—GPT-4o mini outperformed other small models on MMLU tasks; (2) cost-effectiveness—GPT-4o mini costs only 3% of GPT-4o, or 1/3 of GPT-3.5 Turbo.

```
import openai

# {hazard_type} was obtained from the pretrained DL model

# 1. Initialize the API key
openai.api_key = "API_KEY"

def get_hazard_prevention_measures(hazard_type: str) -> str:
    """
# Query ChatGPT-4o-mini for prevention measures given a recognized {hazard_type}.
    """
    # 2. Fill the template
    prompt = (
        f"You are a construction safety expert. "
        f"Based on the provided documents, "
        f"please list specific prevention measures for the hazard: “{hazard_type}.”"
    )

    # 3. Call the ChatCompletion API
    response = openai.ChatCompletion.create(
        model="gpt-4o-mini",
        messages=[
            {"role": "system",
             "content": "You have full access to the TOSHA construction safety laws, regulations, standards,
and specs."},
            {"role": "user", "content": prompt}
        ],
        temperature=0.0,
        max_tokens=300
```

Figure 3: Python code (partial) to call ChatGPT-4o-mini.

3.2.3 Examples of Responses Generated by LLM-RAG

Table 3 presents four examples of responses generated by the ChatGPT-4o-mini API, detailing hazard prevention measures based on TOSHA safety regulations for a specified hazard type.

No.	Image	Hazard Description	Precommended revention Measures	Source Document
1		Hazard: Fall from height Scenario: Gap between scaffold planks and wall, no safety net.	1. Seal gaps with rigid covers ≤3 cm 2. Install guardrails ≥90 cm + mid-rails 35–55 cm + toe-boards ≥10 cm 3. Hang CNS 14252 safety nets ≤7 m drop, 2.5–4 m extension 4. Use full-body harness per CNS 7534/CNS 14253 when nets/rails absent	1. CSHFS #21 2. CSHFS #20 3. CSHFS #22 4. LIA #28
2		Hazard: Fall to lower level Scenario: Worker perched on rebar cage edge without platform or guardrail.	1. Install scaffold planks or secure platforms across rebar cages 2. Erect guardrails ≥90 cm high + mid-rails + toe-boards ≥10 cm on all edges 3. Cap or bend exposed rebar ends and cover with protective caps 4. Use full-body harness per CNS 7534/CNS 14253 when working near open edges	1. CSHFS #56-2 2. CSHFS #20 3. CSHFS #67 4. LIA #28
3		Hazard: Fall Scenario: Workers stand on unstable planks around rebar column, no guardrails or proper platform.	1. Install stable platform/scaffold planks spanning full bay 2. Erect guardrails ≥90 cm + mid-rails + toe-boards ≥10 cm on edges 3. Cap/bend exposed rebar + use protective caps 4. Wear full-body harness anchored to certified lifeline if edges unprotected	1. CSHFS #56-2 2. CSHFS #20 3. CSHFS #67 4. LIA #28
4		Hazard: Fall from height Scenario: Worker perched on narrow beam without platform or harness.	1. Install full-width scaffold planks or work platform 2. Erect guardrails ≥ 90 cm high + mid-rails + toe-boards ≥ 10 cm on all open edges 3. Require full-body harness anchored to certified lifeline when platform or rails absent	1. CSHFS #56-2 2. CSHFS #20 3. LIA #28

Table 3: Examples of generated responses from ChatGPT 4o mini.

3.2.3 Performance Evaluation

To evaluate the effectiveness of the hazard prevention measures recommended by the LLM-RAG model, we applied the performance metrics defined in Subsection 2.3.2—using a five-point Likert scale—to a set of 30 randomly selected hazard images. The results are summarized in Table 4. Among the three metrics assessed, Completeness of Hazard Descriptions achieved the highest mean score and approval rate.

Metric	Hazard Description	Prevention Measures	Source Documents
Mean score	4.2	3.7	3.2

Table 4: Performance metrics of the recommended prevention measures.

The following key observations emerged from the test cases:

- The single most influential factor was the accuracy of hazard identification: when the model misclassified the hazard type, it recommended inappropriate prevention measures and cited incorrect source documents. Currently, these misclassifications stem from the limited visual hazard recognition capabilities of the pre-trained model—improving this is a primary task for enhancing the proposed system.
- Dataset imbalance significantly impacted hazard classification: many images were incorrectly labeled as "H01–Fall & Rolling." Although this hazard is the most common on site, its overrepresentation skewed the model's performance. Future work should focus on collecting more images of other hazard types.
- Hallucination issues were observed in sourcing: the LLM-RAG model sometimes cited incorrect documents even when its prevention measures were accurate. Addressing these hallucinations is essential to improve the model.

4 Conclusions and Contributions

This study represents one of the earliest efforts to integrate a Retrieval-Augmented Generation (RAG)–enhanced large language model (LLM) with deep learning (DL)–based visual hazard recognition for the automated generation of construction hazard prevention measures. Our results demonstrate that the proposed LLM-RAG framework can automatically generate practical, site-specific safety solutions. Such a tool offers construction managers and safety personnel significant benefits, addressing an industry that consistently records the highest number of workplace fatalities. By coupling the LLM-RAG model with pretrained DL detectors and CCTV monitoring, it is possible to achieve continuous, around-the-clock hazard surveillance. When paired with a real-time alert system, this approach supports proactive safety management, bringing us closer to the goal of identifying and eliminating hazards before they occur.

Although the results have demonstrated the feasibility of the proposed model, further efforts are necessary to enable its practical implementation. Future research should focus on the following directions:

- Enhancing visual hazard recognition performance — While image recognition technologies have advanced significantly, applying them to complex construction scenarios such as hazard detection remains challenging. In this study, four pre-trained deep learning models (GoogleNet, Inception-v3, ResNet-101, and DenseNet-201) were used, with accuracy rates plateauing around 80%. Future research should address issues such as dataset imbalance and the identification of compound hazards to improve recognition performance.

- Reducing hallucination of LLM-RAG model—it was found in this research that hallucination effects was serious in retrieving related source safety regulations. This will reduce the credibility of the generated solution significantly. Techniques such as hybrid retrieval, citation-aware generation, Chain-of-reasoning grounding, etc. can be adopted in reducing hallucination effects.
- Automated system to implement the proposed model in practical construction safety management including the real-time safety warning system to inform hazard information the workers themselves and the site safety managers, and a historical lesson-learned system that accumulates hazard prevention experiences are also desirable in developing future proactive safety management strategies.

Acknowledgements

This research project was funded by the National Science and Technology Council, Taiwan, under project No. MOST 111-2221-E-324 -011 -MY3. Sincere appreciations are given to the sponsor by the authors.

References

[1] X. Xia, P. Xiang, S. Khanmohammadi, T. Gao, and M. Arashpour, "Predicting safety accident costs in construction projects using ensemble data-driven models," Journal of Construction Engineering and Management, vol. 150, no. 7, Art. no. 04024054, 2024.

[2] ILO, "Safety and health in the construction sector—Overcoming the challenges," International Labour Organization, Geneva, 2014.

[3] Eurostat, "Accidents at work statistics," Eurostat, 2020. [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Accidents_at_work_statistics. [Accessed: Jun. 8, 2024]

[4] P. Mitropoulos, T. S. Abdelhamid, and G. A. Howell, "Systems model of construction accident causation," Journal of Construction Engineering and Management, vol. 131, no. 7, pp. 816–825, 2005.

[5] A. Perlman, R. Sacks, and R. Barak, "Hazard recognition and risk perception in construction," Safety Science, vol. 64, pp. 13–21, 2014.

[6] A. Albert, M. R. Hallowell, and B. M. Kleiner, "Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study," Journal of Construction Engineering and Management, vol. 140, 2014, doi:10.1061/(ASCE)CO.1943-7862.0000790.

[7] S. Bahn, "Workplace hazard identification and management: The case of an underground mining operation," Safety Science, vol. 57, pp. 129–137, 2013.

[8] G. Carter and S. D. Smith, "Safety hazard identification on construction projects," Journal of Construction Engineering and Management, vol. 132, no. 2, pp. 197–205, 2006.

[9] S. M. J. Uddin, A. Albert, A. Alsharef, B. Pandit, Y. Patil, and C. Nnaji, "Hazard recognition patterns demonstrated by construction workers," International Journal of Environmental Research and Public Health, vol. 17, no. 21, Art. no. 7788, 2020.

[10] M. M. Alateeq, P. P., F. R., and M. A. S. Ali, “Construction site hazards identification using deep learning and computer vision,” Sustainability, vol. 15, no. 3, Art. no. 2358, 2023.
[11] M. Liu, M. Liang, J. Yuan, J. Wang, and P.-C. Liao, “Time lag between visual attention and brain activity in construction fall hazard recognition,” Automation in Construction, vol. 168, Art. no. 105751, 2024.
[12] W. T. Hsiao, W. D. Yu, C. Y. Tang, and A. Bulgakov, “Subtle fall hazard scenario identification using transfer learning,” in Proceedings of the 6th International Conference on Architecture, Construction, Environment, and Hydraulics (ICACEH 2024), Taichung, Taiwan, Dec. 6, 2024, pp. 1–4.
[13] S. Pandithawatta, S. Ahn, R. Rameezdeen, C. W. K. Chow, and N. Gorjian, “Systematic literature review on knowledge-driven approaches for construction safety analysis and accident prevention,” Buildings, vol. 14, no. 11, Art. no. 3403, 2024.
[14] Y. M. Goh and C. U. Ubeynarayana, “Construction accident narrative classification: An evaluation of text mining techniques,” Accident Analysis & Prevention, vol. 108, pp. 122–130, 2017.
[15] B. Esmaeili, M. R. Hallowell, and B. Rajagopalan, “Attribute-based safety risk assessment. II: Predicting safety outcomes using generalized linear models,” Journal of Construction Engineering and Management, vol. 141, no. 8, Art. no. 04015022, 2015.
[16] K. Kang and H. Ryu, “Predicting types of occupational accidents at construction sites in Korea using a random forest model,” Safety Science, vol. 120, pp. 226–236, 2019.
[17] R. Squillante Jr., D. J. Santos Fo, N. Maruyama, F. Junqueira, L. A. Moscato, F. Y. Nakamoto, P. E. Miyagi, and J. Okamoto Jr., “Modeling accident scenarios from databases with missing data: A probabilistic approach for safety-related systems design,” Safety Science, vol. 104, pp. 119–134, 2018.
[18] W.-D. Yu, W.-T. Hsiao, T.-M. Cheng, H.-S. Chiang, and C.-Y. Chang, “Describing construction hazard images identified from site safety surveillance video,” in Proceedings of the 3rd International Civil Engineering and Architecture Conference (CEAC 2023), Kyoto, Japan, Mar. 17–21, 2023, pp. 937–948, 2024. doi:10.1007/978-981-99-6368-3_76
[19] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang, “Retrieval-augmented generation for large language models: A survey,” arXiv preprint arXiv:2312.10997, 2023. doi:10.48550/arXiv.2312.10997
[20] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., “Training language models to follow instructions with human feedback,” Advances in Neural Information Processing Systems, vol. 35, pp. 27730–27744, 2022.
[21] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4700–4708). IEEE. https://doi.org/10.1109/CVPR.2017.465 (Preprint available at: http://arxiv.org/abs/1608.06993)

[22] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[23] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009. doi:10.1016/j.ipm.2009.03.002
[24] Y. Wang, B. Xiao, A. Bouferguene, M. Al-Hussein, H. Li, "Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning," Advanced Engineering Informatics, Vol. 53, 101699, 2022, https://doi.org/10.1016/j.aei.2022.101699.
[25] J. Boone and R. Boone, "Analyzing Likert data," Journal of Extension, vol. 50, no. 2, pp. 1–5, 2012.
[26] W.-T. Hsiao, W. D. Yu, T.-M. Cheng, and A. Bulgakov, "Preliminary study on image captioning for construction hazards," Engineering Proceedings, vol. 74, no. 1, pp. 1–7, 2024. doi:10.3390/engproc2024074020